R E P O R T R E S U M E S

ED 010 990 SE 000 006

THE USE OF INSERVICE PROGRAMS TO DIAGNOSE SOURCES OF RESISTANCE TO INNOVATION.

BY- ROWE, MARY BUDD HURD, PAUL DEHART

PUB DATE 66

EDRS PRICE MF-$0.09 HC-$0.68 17P.

DESCRIPTORS- *CURRICULUM DEVELOPMENT, *ELEMENTARY SCHOOL SCIENCE, *INSERVICE TEACHER EDUCATION, BIOLOGY, INNOVATION, INSERVICE PROGRAMS, PHYSICAL SCIENCES

FACTORS RELATED TO RESISTANCE TO INNOVATION IN THE ELEMENTARY SCHOOL SCIENCE CURRICULUM WERE IDENTIFIED THROUGH THE EVALUATION OF AN INSERVICE EDUCATION PROGRAM. THE SAMPLE WAS COMPOSED OF 290 ELEMENTARY TEACHERS AND ADMINISTRATORS WITH VARIED MATHEMATICS AND SCIENCE BACKGROUNDS, TEACHING EXPERIENCE, AND TEACHING OR ADMINISTRATIVE ASSIGNMENTS. TWICE DURING THE YEAR, THE TEACHERS ATTENDED THREE, 3-HOUR LESSONS IN AN ELEMENTARY SCIENCE INSERVICE PROGRAM. INTEREST AND UTILITY SCALES WERE ADMINISTERED AFTER EACH LESSON AND QUESTIONNAIRES WERE ADMINISTERED AFTER EACH OF THE TWO MAJOR SESSIONS. ALL TEACHERS INDICATED INTEREST IN THE PROGRAM. TEACHERS WITH 16 TO 20 YEARS OF EXPERIENCE WERE NOT AS INTERESTED, NOR DID THEY FIND THE PROGRAM AS USEFUL, AS THE OTHER EXPERIENCE SUBGROUPS. TEACHERS WITH UP TO 15 HOURS OF COLLEGE MATHEMATICS FOUND THE PROGRAM MORE USEFUL THAN TEACHERS WITH MORE HOURS OF MATHEMATICS. THERE WAS NO SIGNIFICANT RELATIONSHIP BETWEEN THE TEACHERS' SCIENCE BACKGROUND, WHETHER THEY HAD TAKEN A SCIENCE METHODS COURSE, AND THEIR RATING OF THE PROGRAM FOR USEFULNESS AND INTEREST. THE TEACHERS INDICATED A BELIEF THAT CHILDREN LEARN BEST BY DOING, BUT EXPRESSED A CONCERN FOR RELATED DISCIPLINE PROBLEMS. ADMINISTRATORS WERE MORE CONCERNED WITH THE TEACHER'S LACK OF KNOWLEDGE OF SCIENCE CONTENT THAN WITH DISCIPLINE PROBLEMS ARISING FROM STUDENT ACTIVITIES. THIS ARTICLE IS PUBLISHED IN THE "JOURNAL OF RESEARCH IN SCIENCE TEACHING," VOLUME 4, ISSUE 1, 1966. (AG)

AUG - 4 Rec'd

ED010990

JOURNAL OF RESEARCH IN SCIENCE TEACHING

Volume 4 — Issue 1 — 1966

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

CONTENTS

- From the Editor 1
- The Use of Inservice Programs to Diagnose Sources of Resistance to Innovation: *Mary Budd Rowe and Paul DeHart Hurd* 3
- Three Emphases in Teaching Biology—A Statistical Comparison of Results: *Robert E. Yager and John W. Wick* 16
- Inquiry Training and Problem Solving in Elementary School Children: *David P. Butts and Howard L. Jones* 21
- The Relationship of Three Factors in Printed Materials to Student Achievement: *Edward J. Eaton* 28
- Research Reports
  - Trends in Assessing Teacher Effectiveness: *Gladys S. Kleinman* 37
  - The Influence of Teacher Attitudes and Individual Differences on Pupil Achievement with Programed Science Materials: *Alton L. Taylor* 38
  - Using Programed Instruction to Teach High School Chemistry: *J. Henry Sayles* 40
  - The Development of Concepts of Space, Matter, and Energy in Students at the College Level: *Leona K. Adler* 41
  - Earth Science Gains Momentum in the Secondary Schools: *Robert E. Boyer and John L. Snyder* 43
- The Concept of "Model" and Educational Research: *Henry Robert Weinstock* 45
- Analyzing Test Responses with Symbolic Logic: *Gary R. Smith* 52
- Preparing the Junior High School Science Teacher, Model 1980: *Matthew H. Bruce, Jr.* 59
- Books Received 63

Published under the auspices of
NATIONAL ASSOCIATION FOR RESEARCH IN SCIENCE TEACHING *and*
ASSOCIATION FOR THE EDUCATION OF TEACHERS IN SCIENCE *by*
John Wiley & Sons, Inc., New York • London • Sydney

SE 000 006

JOURNAL OF RESEARCH IN SCIENCE TEACHING

Editor

J. STANLEY MARSHALL

Department of Science Education

The Florida State University, Tallahassee, Florida

Editorial Advisory Board

J. DARRELL BARNARD
New York University
New York, New York

SIDNEY ROSEN
University of Illinois
Urbana, Illinois

WILLIAM COOLEY
Harvard University
Cambridge, Massachusetts

IRWIN L. SLESNICK
Western Washington State College
Bellingham, Washington

PAUL deH. HURD
Stanford University
Palo Alto, California

HERBERT A. SMITH
Colorado State University
Fort Collins, Colorado

JOSEPH F. JORDAN
John Wiley & Sons, Inc.
New York, New York

FRANK X. SUTMAN
Temple University
Philadelphia, Pennsylvania

EUGENE C. LEE
Emory University
Atlanta, Georgia

STEPHEN S. WINTER
University of Buffalo
Buffalo, New York

JOHN MASON
Michigan State University
East Lansing, Michigan

ROBERT E. YAGER
University of Iowa
Iowa City, Iowa

Published quarterly (March, June, September, December) by John Wiley & Sons, Inc. covering one volume annually. © Copyright 1965 by National Association for Research in Science Teaching. Publication Office at 20th and Northampton Streets, Easton, Pennsylvania 18043. Executive, Editorial, and Circulation Offices at 605 Third Avenue, New York, New York 10016. Application for Second-class postage pending at Easton, Pa. Subscription price, $10.00 per volume. Single issue price, $3.00. Foreign postage, $0.50 per volume. Manuscripts should be submitted in duplicate to Prof. J. Stanley Marshall, Department of Science Education, The Florida State University, Tallahassee, Florida 32306.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY H. Craig Sipe, Editor

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

SEP 2 6 1966

JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 4, PP. 3–15 (1966)

A novel approach to the evaluation of inservice programs in the propagation of change in elementary school science—one which employs techniques of engineering research in locating sources of resistance.

The Use of Inservice Programs to Diagnose Sources of Resistance to Innovation*

MARY BUDD ROWE† and PAUL DEHART HURD

School of Education, Stanford University, Stanford, California

Introduction

This paper reports on a study conducted to determine what factors should be considered in designing an efficient program for diffusing new styles of teaching elementary school science. It suggests how evaluation of inservice programs can serve a diagnostic function for consultants confronted with the task of getting teacher innovations into general practice. Finally, it offers an argument for the use of engineering methodology in analyzing data associated with programs designed to produce change.

In comparison with the complexity of the task of diffusing a new curriculum, curriculum development in elementary science is a relatively simple process. Every summer, curriculum groups write or revise new elementary science courses. Every fall, a few schools try out the new programs. Most schools go on with their usual routines unaware of new developments.

A new curriculum is a plan on paper that must somehow get translated into an action program. The translating device is the inservice program. But inservice education programs aimed at introducing innovations in elementary science teaching must, by the very nature of the innovation, disturb the quiescent state of a complex system of human, institutional, and environmental variables. The sources of resistance to change are far more diverse than the simple complaint that teachers do not know enough science. Those inservice programs, for example, that attempt to convey something of the style of teaching demanded by the new curricula in science present especially difficult problems for school personnel.

This study reports one example of how consultants may employ inservice programs to diagnose sources of resistance to innovation in teaching elementary science. Based on such diagnoses, consultants may then develop strategies for minimizing resistance, or, to put it in other terms, they may discover how to persuade school personnel to assume an experimental attitude toward "new" elementary science curricula.

Inservice programs are conceptualized by the investigators as strategies designed to get components of a complex social institution to interact in such a way as to achieve a designated purpose, namely, innovation. Evaluation of inservice programs, therefore, falls into the category of engineering research since, to pertain to practical goals, the evaluation must be conducted under and be relevant to operating conditions of the system. The consultant disturbs the "steady state" of the school system in some defined manner that he predicts or forecasts will move the system toward the new, innovative state. He then seeks to discover in what state the system is *after* his effort

* Support for this research was provided by the Shell Companies Foundation, Incorporated, and the Procter and Gamble Faculty Research Fund.

† Current address: Teachers College, Columbia University, New York, N. Y.

TABLE I-A
Description of the Population

Site	No. of districts	No. of participants			Semester hr. of science and/or mathematics								Science methods course[a]		
					Science				Math.				Lab	No[b] lab	
		Males	Females	Total	0–4	5–10	11–16	More	0–4	5–10	11–16	More	+	−	Total
01	4	9	20	29	6	8	9	6	17	10	1	1	3	15	18
02	1	9	16	25	3	4	5	12	7	5	9	3	1	13	14
03	10	13	16	29	3	6	5	11	15	5	2	4	1	9	10
04	4	7	13	20	0	4	8	8	8	6	5	1	0	12	12
05	6	11	10	21	1	7	3	10	13	6	1	1	2	11	13
06	5	8	21	29	6	13	5	5	17	7	3	2	2	11	13
07	3	11	16	27	5	7	8	7	17	7	3	3	0	12	12
08	9	13	10	23	2	9	6	6	19	4	0	0	2	10	12
09	7	7	24	31	7	9	8	7	21	6	2	2	1	16	17
10[c]	5	10	23	33	6	15	7	5	19	9	3	1	1	23	24
11[d]	7	17	6	23	3	7	6	4	6	13	1	0	2	12	14
Totals	61	115	175	290	42	89	70	81	159	78	29	18	15	144	159
% of *N*		40	60												55

[a] Number who had science methods course prior to the Mobilab program or concurrently with it.
[b] (+) laboratory experience with methods course; (—) no laboratory experience.
[c] 6 hr less instruction.
[d] Session restricted to administrators.

TABLE I-B
Description of the Population—Reported Years of Teaching Experience

Site	Years of experience					Teaching assignments			
	1–5	6–10	11–15	16–20	21 or more	K–3	4–6	7–9	Adm.[a]
01	7	9	3	4	5	7	18	4	0
02	3	6	2	2	10	4	13	4	4
03	8	7	7	2	3	2	17	9	1
04	5	5	8	0	2	4	9	7	0
05	8	6	2	3	2	5	8	8	0
06	4	7	5	2	11	8	17	1	3
07	12	2	7	2	4	5	19	1	2
08	8	3	7	2	3	2	13	7	1
09	10	3	6	6	6	7	21	1	2
10	12	8	5	1	6	8	22	1	2
11	3	5	8	2	2	—	—	—	23
Total *N* per category	81	61	60	26	54	52	157	43	38
% of *N*						18	54	15	13

[a] Administrators (three also have some teaching responsibility).

and by the use of *post hoc* hypotheses to connect forecaster and criterial terms. Forecasting furnishes a direct and immediate basis for action by the consultant provided that he represents in the data whatever sources of variability may be operating within the criterial condition. The function of the engineering technique is to tell the consultant how to operate on the system to cause it to move toward specified objectives, and thus to employ statistical estimation rather than hypothesis testing techniques.

Program Description

In 1960 the Colorado Department of Education introduced a portable inservice education program in elementary science. The department placed in service a fully equipped teaching laboratory housed in a 40-ft. trailer. At each of the eleven sites to which the trailer moved, a science consultant from the Colorado Department of Education conducted an inservice program in science and mathematics for elementary teachers. Table I describes the teacher population that took part in the inservice program.

The Mobile Laboratory visited each site twice in the year and stayed for three days at all sites except one. The first visit to a site consisted of three, 3-hr. lessons of a "learn-by-doing" approach to the life sciences. The second visit of the same duration centered around a "learn-by-doing" physical science sequence chosen to develop and use measurement and number concepts.

Procedure

First Level Indicators of Resistance

To be economical in achieving his purpose of changing the resistance of school personnel to innovation, the consultant must specify the target groups he wishes to reach through an inservice program. He then needs to discover how wide a target in terms of teacher experience, course background of participants, and grade level interests he can serve satisfactorily with one program.

At the end of each session participants evaluated the session on two, 3-part scales: *Not Interesting; Interesting; Very Interesting;* and *Not Useful, Useful, Very Useful.* No finer scale discriminations seemed appropriate in view of the basic intent, namely, to detect sources of resistance, not to measure how much resistance occurred. People checking the low and middle categories of the two scales represent potentially resistant targets. Since the consultant needs to detect sources of resistance, it is valuable in the analysis of the "interest-use" data provided by the rating scales to collapse the two lower categories into one (0,1). (See Table III.) The argument for this strategy stems from the assumption that teachers are especially prone to avoid making negative remarks; thus a teacher would be likely to check a middle category on each scale unless very strong feelings or beliefs were involved. Checks in the lowest categories of *interest* and *use* seldom appeared, thus supporting the assumption.

By collapsing the *interest–use* scales one can then regard people as belonging to one of two groups, potential *resistors* (0,1) and potential *acceptors* (2). Using these scales as crude "state of the system" indicators, the population is then partitioned in several ways to determine whether the interest–use ratings are significantly sensitive to the partitions. The first step in analysis is to discover to what extent interest and use are correlated. How likely, for example, is a teacher to check a session as "very useful," given that she rated a session very "interesting," or how likely is she to check "very useful" when she called a session "not very interesting"? Table II indicates that a "very useful" response is 50% more likely to occur when a very interesting response occurs than when a "not very interesting" response occurs. In short, a high "impact" program tends to increase the chances that the program will also be perceived as useful.

Of the two ratings, *usefulness* appears to be the more sensitive indicator of possible sources of resistance to change. Interest

stays high over all partitions and thus is not a useful indicator of resistance. When the population was partitioned first according to grade level (Table III), then by years of teaching experience (Table IV), and finally by background in mathematics (Table V) and science, the "interest" relation remained insensitive to the partition, but the usefulness relation proved to be sensitive on all except the science partition. Examination of the body of Tables III through V suggests that interest is high no matter how one

TABLE II
Extent to Which Perceived Interest and Usefulness Are Correlated for Each of Six Days and for the Program as a Whole

	First visit			Second visit			
Day	1	2	3	4	5	6	Overall
Estimated correlation coefficient r	0.46	0.56	0.23	0.52	0.50	0.39	0.50
χ^2 Test that $p = 0$	Highly sig.	Highly sig.	Highly sig.	Highly sig.	Highly sig.	Highly sig.	Highly sig.
95% confidence interval for rating "very useful" given "very interesting"	0.76 ± 0.08	0.60 ± 0.10	0.78 ± 0.07	0.60 ± 0.10	0.57 ± 0.08	0.77 ± 0.08	0.70 ± 0.04
95% confidence interval for rating "very useful" given "not very interesting"	0.32 ± 0.13	0.08 ± 0.10	0.20 ± 0.31	0.10 ± 0.12	0.06 ± 0.14	0.30 ± 0.17	0.15 ± 0.06

Note: The confidence intervals indicate that a "very useful" response is 50% more likely to occur when a "very interesting" response has occurred than when a "not very interesting" response has occurred.

TABLES III-A and III-B
Relation of Current Teaching Level to Perceived Interest (A) and Perceived Usefulness (B) of the Program, All Sessions Combined

	A. Interest					B. Usefulness				
	Interest rating		Total		$n_i(\hat{p}_i - \hat{p})^2$,	Usefulness rating		Total		$n_i(\hat{p}_i - \hat{p})^2$,
Grade level	0,1	2	m_i	$\hat{p}_i$	$\hat{p} = 0.735$	0,1	2	μ_i	$\hat{p}_i$	$\hat{p} = 555$
K–4	60	168	228	0.737	0.00091	110	109	219	0.498	0.712
4–6	173	502	675	0.744	0.05460	301	348	649	0.536	0.234
7–9	58	137	195	0.702	0.187	75	115	190	0.605	0.475
Administration	37	101	138	0.732	0.00124	42	19	133	0.684	2.22

(A) $\chi_3^2 = 1.23$; $\alpha = 0.75$; not significant.
(B) $\chi_3^2 = 14.8$; $\alpha = 0.02$; significant.

Note: Table III-B shows that administrators and primary teachers are the main contributors to the finding that perceived usefulness of the program seems to be influenced by current teaching level. Note, however, that the primary teachers contribute a high proportion of 0,1 ratings while the administrators choose proportionately more 2 ratings.

TABLES IV-A and IV-B
Relation of Years of Teaching Experience to Perceived Interest (A) and Usefulness (B) of the Program, All Sessions Combined

	A. Interest					B. Usefulness				
	Interest rating		Total			Usefulness rating		Total		
Years of experience	0,1	2	n_i	$\hat{p}_i$	$n_i(\hat{p} - \hat{p})^2$, $\hat{p} = 0.736$	0,1	2	n_i	$\hat{p}_i$	$n_i(\hat{p}_i - \hat{p})^2$, $\hat{p} = 0.555$
1–5	76	226	302	0.749	0.051	124	174	298	0.584	0.252
6–10	65	180	245	0.735	0.000	110	131	241	0.543	0.035
11–15	73	203	275	0.735	0.000	117	144	251	0.573	0.082
16–20	51	116	166	0.700	0.216	90	65	155	0.420	2.825
21 up	56	167	223	0.750	0.04[illegible]	85	130	215	0.604	0.519

(A) $\chi_4^2 = 1.60$; $\alpha = 0.81$; not significant.
(B) $\chi_4^2 = 15.0$; $\alpha = 0.005$; significant (main cause is the 16–20 year group).

TABLES V-A and V-B
Relation of Number of Semester Units of Mathematics to Perceived Interest (A) and Usefulness (B) of the Program, All Sessions Combined

	A. Interest					B. Usefulness				
	Interest rating		Total			Usefulness rating		Total		
Number of math units	0,1	2	n_i	$\hat{p}_i$	$n_i(\hat{p}_i - \hat{p})^2$, $\hat{p} = 0.731$	0,1	2	n_i	$\hat{p}_i$	$n_i(\hat{p}_i - \hat{p})^2$, $\hat{p} = 0.556$
0–4	185	504	689	0.732	0.001	313	350	663	0.527	0.560
5–10	93	237	330	0.718	0.059	125	189	314	0.601	0.635
11–15	29	91	120	0.758	0.088	45	82	127	0.645	1.010
16 or more	18	52	70	0.743	0.010	35	27	62	0.435	0.910

(A) $\chi_3^2 = 0.80$; $\alpha = 0.85$; not significant.
(B) $\chi_3^2 = 12.6$; $\alpha = 0.005$; significant (caused by all groups).

partitions the population. The tables suggest, however, that a high interest program is not enough, the program must be perceived as *useful* before teachers will make any commitment to change. But not all groups perceived the program to be useful to the same extent, as Tables III-A and -B illustrate. Administrators and primary teachers are the main contributors to the finding of a significant relation between current role assignment and perceived "usefulness" of the inservice program, although not for the same reason. Primary teachers choose the 0,1 or lower ratings on the "usefulness" scale about as often as the 2 rating ($p_i = 0.498$). Administrators, on the other hand select the 2 or higher rating more than twice as often as they choose the 0,1 category ($p_i = 0.684$). Data in the body of Table III-B suggest that the program may be "off course" for the primary group.

Another point of interest that piques one's curiosity is the finding, so far unexplained, reported in Table IV-B that teachers with 16–20 yr. of experience are probably a particular source of resistance. They are neither as interested as other groups nor do they, in general, find the program as useful. The 16–20-yr. group sorts out fairly consistently throughout the study as differing in response pattern from other experience categories.

In the design of inservice programs, the consultant needs to know how the academic background of teachers bears on their perception of the usefulness of a program. Is the program rated on *interest* and *usefulness* in the same way by teachers with a great deal

of science and mathematics as by teachers with little experience in college science and mathematics? You will observe from the population description in Table I that relatively few of the participants had taken much mathematics in college. *Over half the group had never had any college mathematics.* For people with up to 15 semester hours of mathematics the program was both interesting and useful. People with 16 or more hours in mathematics found the program interesting but not particularly useful. Thus the amount of mathematics a person had appears to describe one parameter of the target.

By way of contrast, there appears to be no significant relation between the number of semester hours in science a person had taken and his rating of the program on either *interest* or *use*. Supporting data from content analysis make it safe to interpret the stability of the usefulness rating in this partition to mean that the ideas presented and the style of representing them probably differed markedly from college science courses to which participants had been exposed. Probably the same interpretation accounts for a failure to find a significant relation between "interest" or "use" ratings and whether teachers had taken a science methods course.

Content Analysis

Inservice of the type employed in the Colorado Mobile Laboratory Program calls into play, and possibly into conflict, elements of the teacher's value or belief system by furnishing a model against which the teacher checks his practices and beliefs. Through content analyses the following propositions were evaluated: (*1*) *If it appears to the teacher that the model for teaching science in the new way is acceptable in principle but its implementation is perceived to risk the quality of the teaching act, then the teacher will resist making the change.* That is, teachers regard themselves as having "expertise" in the teaching act itself and they tend to reject new practices, *even when these practices are consonant with their beliefs,* if they think that their expertise rating is likely to suffer. Thus, if a new program appears to present potential management or discipline problems, as new science curricula do, teachers will tend to reject that program. (*2*) *Administrators, on the other hand, perceive the problems of teaching elementary science as stemming less from environmental and classroom management (discipline) factors and more from lack of teacher know-how in science.* (*3*) *If teachers are correct in assessing the principal's norms or student behavior in the classroom, then there should be no difference in the importance each group attaches to management or discipline.* Tables VI-A and VI-B illustrate the relation of role to perception of difficulties that would be faced in teaching new science programs. They tend to support the propositions and indicate to the consultant that he may need to

TABLES VI-A and VI-B
Relation between Role and Difficulties Perceived in Teaching Science

	A. Know how			B. Management		
	$\hat{p}_i$	n_i	$n_i(\hat{p}_i - \hat{p})^2$	$\hat{p}_i$	n_i	$n_i(\hat{p}_i - \hat{p})^2$
Teachers	0.137	240	0.312	0.347	240	0.054
Administrators	0.406	37	2.01	0.244	37	0.287

$\hat{p} = 0.173$; $\chi_1^2 = [2.322/(0.173)(0.837)] = 16.2$; highly significant.
$\hat{p} = 0.332$; $\chi_1^2 = [0.341/(0.332)(0.668)] = 1.53$; not significant.

(A) "Know how" includes all those statements which express a lack of appropriate knowledge to teach science, e.g., "I don't know enough," "I lack confidence," "I don't know how."

(B) "Management" includes statements which express concerns related to discipline, e.g., "children get too excited," "keeping everyone working is a problem." "Some might get into trouble," "Size of class is too big to teach this way."

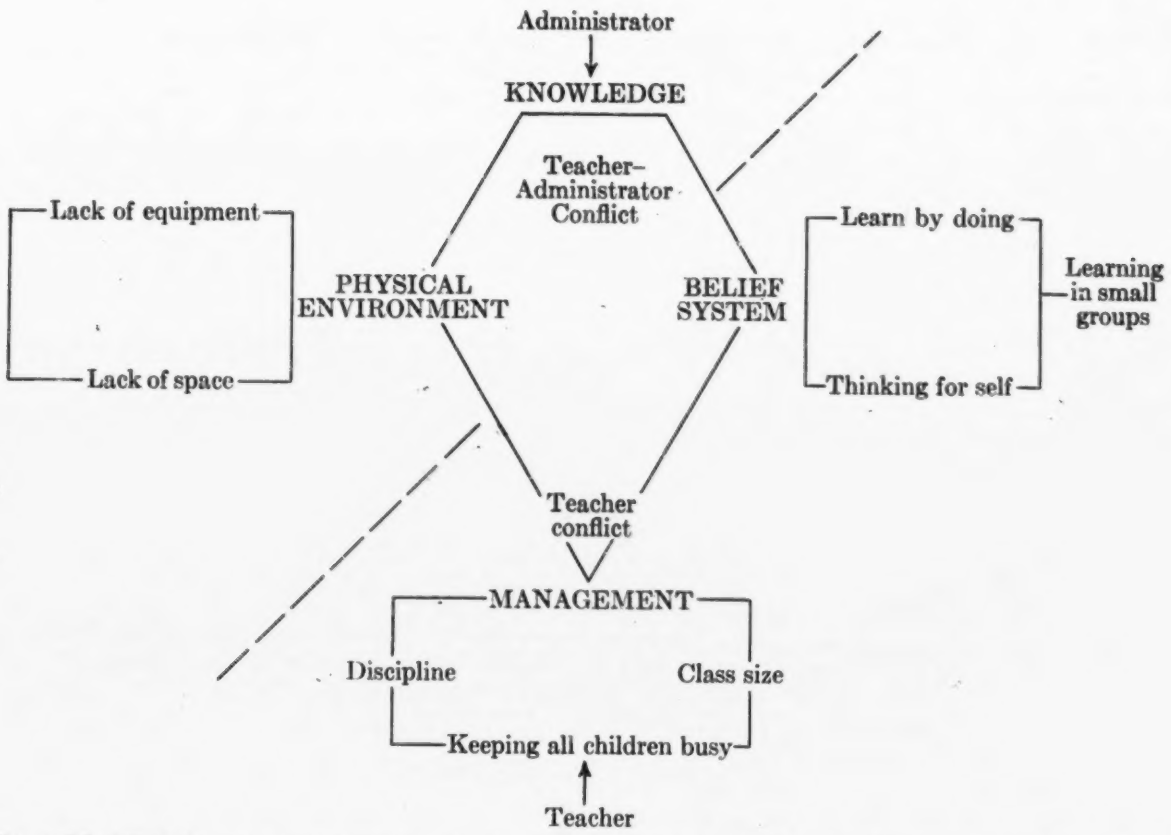

Fig. 1. Model of the set of relations examined through content analysis. Teachers focus primarily on factors lying beneath the diagonal while administrators concentrate mainly on factors above the diagonal.

engineer routes for resolving role differences in the perception of difficulties innovation poses.

Teachers responded to open-ended questions administered at the conclusion of each visit. The questionnaires asked such things as the following: What advantages would there be in teaching a program like this to your students? What difficulties would you face if you tried to introduce a program like this in your school? What suggestions would you make to colleges planning science methods courses for elementary teachers? Responses to these and other questions were coded using content categories derived from the three assumptions described above.

Figure 1 shows the model we wished to evaluate. Tables VII and VIII show some of the relations indicated by the model. Conflict aroused by the inservice program has to do mainly with the relation of elements in the belief system to class discipline or management factors. Teachers expressed belief that children learn best by doing. On the other hand, teachers said, "Some children get too excited and they can't settle down," "Some think science is magic," "Some might get bored," "Keeping everyone busy is a problem." The administrator's perception of discipline enters teachers' rationalizations. Administrators, according to teachers, want quiet, orderly classrooms. If children make noise, move around, and appear to be out of order, as they might in several of the new science programs, teachers say the administrator would censure them. Administrators, on the other hand, rarely mention discipline or management problems. Instead they are more prone to think that teachers lack knowledge of science content than teachers are prone to think it about themselves. The relation between the role people occupy and the frequency with which they express the feeling that they do not have enough knowledge to teach science is

TABLE VII
Association between Elements of the Belief System, Problems of Management, Availability of Equipment and/or Space

	E_{a+b} vs. G_b Physical factors vs. management	E_{a+b} vs. C_{a+b} Physical factors vs. learn-by-doing	G_b vs. C_{a+b} Management vs. learn-by-doing	G_b vs. A_b Management vs. group factors
Correlation	0.09	0.10	0.16	0.15
$\chi^{(1)}$ value	2.5; not sig.	3.0; not sig.	7.2; 0.01 sig.	6.9; 0.01 sig.
$\hat{p} = n_{11}/n_1$: (estimate of p)	0.38	0.40	0.48	0.22
95% confidence interval for p	0.38 ± 0.07	0.04 ± 0.07	0.48 ± 0.10	0.22 ± 0.10
$\varphi = n_1/n$:	0.36	0.37	0.37	0.15

(E_{a+b}) Availability of equipment and space.
(G_b) Problems of classroom management.
(C_{a+b}) Belief in "thinking for self" and/or "learning-by-doing."
(A_b) Expressed approval of small group experiences.

significant ($\alpha = 0.005$). The main contributor to the finding is the administrator group (see Table VIII). They think teachers don't know enough.

As a check on the apparent sensitivity of the usefulness scale to teaching assignment reported in Table III, the association between grade level taught and perceived relevance ($A_c - A_b$) of the program was examined. To be consistent, the relationship between grade level and relevance shown in Table IX should be significant, and probably can be counted so ($\alpha = 0.06$). The table confirms the supposition made earlier that the program did not hit all grade levels with equal effectiveness. Special modifications to accommodate the primary teachers are necessary. This group does not perceive the content of the program to be as relevant to its pupils as other groups do. Neither, as it turns out, does it think there is sufficient time in the school schedule to teach science. Consultants charged with starting new science programs in the primary grades will probably need to design special inservice sessions for primary teachers.

TABLE VIII
Relation of Know How to Grade Level Assignment

Level	0	X_i	$0.173\ n_i = a_i$	$(X_i - a_i)^2/a_i$
1–3	41	9	8.7	0.1
4–6	126	21	25.4	0.8
7–8	40	3	7.5	2.7
Administrators	22	15	−6.4	11.5

$\chi_3^2 = 15.1$; $\alpha > 0.005$ (significant).
Note: The trend in the last column for concern about lack of know how to increase with teaching level.

TABLE IX
Relation of Perceived Relevance of the Program ($A_c + A_d$) to Grade Level Assignment

Level	0	n_1	$\hat{p}_1$	$n_i(\hat{p}_i - \hat{p})^2$, $\hat{p} = 0.30$
1–3	28	22	0.44	0.98
4–6	106	36	0.25	0.35
7 8	27	15	0.36	0.15
Administrators	28	9	0.24	0.13

$\chi_3^2 = [1.61/(0.3)(0.7)] = 7.7$; $\alpha = 0.06$.
Relevance ($A_c + A_d$) refers to two content categories. A_c includes remarks which note that program content of skills are directly applicable to the teachers' work, e.g., "lessons are down to earth," "I could teach this tomorrow," "this way would save time."
A_d includes remarks on the direct applicability of methodological skills learned during the inservice, e.g., "preparation of own demonstrations," "learning how to conduct a field trip near school," "use of microprojector," "experiences with use of simple equipment."

The Inservice Paradox

Inservice programs can manufacture difficulties for teachers when the teaching model

TABLE X-A
Summary of Hypotheses Tested, All Sessions Combined

	Time in schedule	Time to prepare	Know how	Management	Motivation or interest	Learn by doing
Grade level: K–3, 4–6, 4–6, 7–8, adm.	$\chi_3^2 = 10.6$ $\alpha = 0.02$ sig.	—	$\chi_3^2 = 15.1$ $\alpha > 0.005$ Sig.	$\chi_3^2 = 2.67$ Not sig.	$\chi_3^2 = 7.0$ $\alpha = 0.08$ Sig.	$\chi_3^2 = 2.6$ Not sig.
Years teaching; 1–5, 6–10, 11–15, 16–20, 21 plus	—	$\chi_4^2 = 9.0$ $\alpha = 0.06$ Possibly sig.	$\chi_4^2 = 1.9$ Not sig.	$\chi_4^2 = 4.9$ Not sig.	$\chi_4^2 = 11.9$ $\alpha = 0.02$ Sig.	$\chi_4^2 = 6.8$
Units in science: 0–4, 5–10, 11–15, 16 plus	—	$\chi_4^2 = 2.30$ Not sig.	$\chi_4^2 = 5.0$ Not sig.	$\chi_4^2 = 2.4$ Not sig.	$\chi_4^2 = 1.0$ Not sig.	—
Units in math: 0–4, 5–10, 11–15, 16 plus	—	$\chi_3^2 = 1.39$ Not sig.	$\chi_3^2 = 0.95$ Not sig.	$\chi_3^2 = 2.45$ Not sig.	$\chi_3^2 = 2.0$ Not sig.	—
Location: sites 1–11	$\chi_{10}^2 = 8.52$ Not sig.	—	—	$\chi_{10}^2 = 21.4$ $\alpha = 0.02$ Sig.	$\chi_{10}^2 = 8.2$ Not sig.	$\chi_{10}^2 = 18.0$ $\alpha = 0.06$ Sig.

TABLE X-B
Summary of Hypotheses Tested, All Sessions Combined

	Small groups	Relevance to students	Lack of space and equipment	Mode of presentation	Problems expected
Grade level: K–3, 4–6, 7–8, adm.	$\chi_3^2 = 2.41$ Not sig.	$\chi_3^2 = 7.7$ $a = 0.06$ Possibly sig.	$\chi_3^2 = 5.8$ Not sig.	$\chi_3^2 = 10.6$ $a = 0.02$ Sig.	$\chi_3^2 = 9.01$ $a = 0.03$
Years teaching: 1–5, 6–10, 11–15, 16–20, 21 plus	$\chi_3^2 = 0.98$ Not sig.	—	$\chi_4^2 = 8.8$ $a = 0.01$ Possibly sig.	$\chi_4^2 = 4.64$ Not sig.	—
Units in science	—	—	—	—	—
Units in math.	—	—	—	—	—
Location: sites 1–11	$\chi_3^2 = 6.10$ Not sig.	$\chi_{10}^2 = 12.0$ Not sig.	$\chi_{10}^2 = 16.8$ $a = 0.10$ Possibly sig.	—	$\chi_{10}^2 = 6.10$ Not sig.

presented in the programs differs markedly from the model currently in use by teachers. For example, on a pre-session questionnaire, participants listed difficulties they would face in teaching science. The mean number of difficulties listed was 0.7. That is, as long as science is a reading program, only lack of time bothers teachers. At the end of the first visit, the mean number of difficulties per person rose to 4.2. Problems of class management and limited space began to appear. At the conclusion of the second

visit, the mean number of difficulties anticipated, if such a program were introduced, expanded to 5.4.

Although we would normally regard high involvement by children as desirable, how to control that enthusiasm presented an enormous problem to participants. How, for example, to shift from science to another subject or how to keep children quiet in the face of their enthusiasm became an increasing concern as participants tried out lessons with their classes. There is one saving feature to this paradoxical situation. As the sessions progress, the number of perceived advantages to teaching science according to the model provided by the program also rises, so that some base forms for confronting the problems.

Summary

Attention of consultants interested in inservice education is invited to further consideration of product-oriented research as a way of monitoring the success of inservice programs in the propagation of innovation in elementary school science. As useful as information is about the state of a school system with respect to innovation, we rarely take advantage of the inservice program as an indicator of the system's condition. The system tends to resist change and it is the consultant's task to diagnose the sources of resistance. Sometimes the resistance has roots in the inservice program itself. At the same time, resistance to change may be rooted in institutional variables. Diagnosis, prescription, and follow-up treatment appropriate to each district or school system may profitably be based on an engineering rather than a theoretical format for the research design. Engineering research is practical research done to provide immediate information about the state of a system. Generalizability of engineering studies is limited to operational systems like the one reported in this paper. Control, not explanation, is the goal of this kind of assessment. Thus, a large number of propositions are tested and the pattern of findings evaluated (e.g., see Table X). In the data reported here, for example, one learns that primary teachers are not as well served by the program as are all other groups. Similarly, teachers with more than 16 hr. of college mathematics probably require some modification of the program presented to them. The number of credits in science, on the other hand, need not be a factor forcing redesign of the program.

The inservice data suggest, in addition, that teachers and principals not only focus on different difficulties that innovation presents; they often hold conflicting views. The consultant can examine the content of the conflicts and perhaps find a means to reduce them. Faced with such representative findings, the consultant lists alternative actions to take with respect to each group, evaluates the alternatives, selects and tries most likely prescriptions (i.e., chooses forecaster items), institutes the prescriptions, and examines their effects (i.e., compares forecaster and criterial items). In this way he maintains a constant monitoring of the inservice program as it relates to the school system and thus provides himself with a basis for designing corrective procedures to maintain the system "on target."

Research Notes

Courses in educational statistics tend to exclude estimation techniques despite the fact that education is an applied field where control is the goal. Decisions about what to do and when to do it should rest on systems analyses which, in turn, employ statistical estimation. The restricted repertoire of statistical techniques which the lay statistician possesses (and most science educators fall into that category) often stunts his conceptualization of a study because he cannot think how to manage the data. Rather than hamper his work, the science educator would do well to formulate his research questions with abandon, albeit clearly, find an inventive statistician, and form a team. If the questions are interesting and the statistician creative, both members of the team

may find the union profitable. In the description of statistical methods given below, for example, Marshall Sylvan and Carl Morris of Stanford University began with assumptions consistent with the demands of the investigation, derived appropriate statistical treatments for the data, and ended up, to their surprise, with a standard chi square technique for a $2 \times K$ table. The assumptions with which they began are not those usually accompanying the chi-square for an $M \times K$ table. Whether the new assumptions would work for an $M \times K$ table is untested at this time.

Problem 1

Given: K independent binomials: $X_i \sim b(n_i,p_i) 1 \leq i \leq k$ meaning that each X_i has a binomial distribution where n_i observations were made, each observation being 0 or 1 and $p_i = Pr(1)$. X_i is the number of 1's appearing in this sampling situation corresponding to the ith sample.

To Test the Hypothesis:

$H_0: p_i = p_i$ for every $i, j, 1 \leq i \leq k, 1 \leq j \leq k$ against all possible alternatives

$H: p_i \neq p_j$ for some i, j.

There is no known method for testing this hypothesis from the given distribution assumptions. However, approximate tests of this hypothesis may be developed when all the n_i, $1 \leq i \leq k$, are sufficiently large, and when the following assumptions are justifiable:

Assumption I: The X_i, $1 \leq i \leq k$, have a normal distribution, approximately. Then since the mean of a binomial X_i is n_ip_i and the variance is $n_ip_i(1 - p_i)$, we may assume that $X_i \sim N(n_ip_i, n_ip_i(1 - p_i)$, and are independent.

Assumption II: The X_i, $1 \leq i \leq k$, have an approximate Poisson distribution, independently, with X_i having Poisson mean n_ip_i.

Assumption I is justifiable when p_i and $1 - p_i$ are not too close to zero and n_i is large. The well-known central limit theorem then applies to assure us that as $n_i \rightarrow \infty$, $(X_i - EX_i)/(\text{var } X_i)^{1/2}$ has an approximate $N(0,1)$ distribution. For many practical purposes, X_i may be considered normal when n_ip_i and $n_i(1 - p_i)$ both exceed 5, as in this case.

Assumption II is justifiable when n_i is large and p_i is small. Then the well-known Poisson limit theorem for binomials applies and states that if $n_i \rightarrow \infty$ while $n_ip_i \rightarrow \lambda_i$ a fixed, non-zero constant, then the limiting distribution of X_i is Poisson with parameter λ_i.

A Method for Testing the Hypothesis under Assumption I:

$$X_3 \sim N[n_ip_i, n_ip_i(1 - p_i)]$$

A serious difficulty still remains. While much is known about testing for equality of means of a normal random variable, little is known about this problem when the covariance matrix is unknown and, as in this case, functionally dependent upon the unknown mean. A method for eliminating this difficulty suggests itself however. Since n_i is considered to be large, the estimates of the unknown variances are so accurate that in fact they may be assumed known in the sense that the distribution theory of the resulting statistic differs little whether the unknown variance is assumed known and non-random and estimated by the sample variance, or whether the unknown variance is considered random and used. (For example, the t-distribution is nearly normal when the number of degrees of freedom is large.) If, furthermore, p_i is estimated under the null hypothesis, then all of the X_i, $1 \leq i \leq k$, may be used in estimation, thus providing a very good estimate under the null hypothesis. If the variances are still estimated under the alternative hypothesis (H_1 true), then the estimate is known to be good provided the alternative does not differ by a very large amount from the null hypothesis. When H_1 differs greatly from H_0, this estimate might be relatively poor, but in this case the power of the test will be so great that no difficulty can arise in detecting which hypothesis is true.

The two assumptions are commonly made in mathematical statistics. In tests where samples are large, as in this case, they are justifiable.

We now assume

$$Y_i = \frac{X_i}{n_i} \sim N\left(p_i, \frac{p_i(1-p_i)}{n_i}\right)$$

and assume that $\sigma_i^2 = P_i(1-p_i)/n_i$ may be considered known (even though p_i is not).

Hence it is desired to test $H_0: p_i = p_j$ for all i, j, when $Y_i \sim N(p_i, \sigma_i^2)$.

The maximum likelihood estimate of p_i under the alternative hypothesis is $Y_i = X_i/n_i$ and under the null hypothesis when all $p_i = p$ is obtained by maximizing

$$\frac{1}{(2\P)^{K/2}\sigma_1 \ldots \sigma_k}\left(\exp\left[-\frac{1}{2}\sum\frac{(Y_i-p)^2}{\sigma_i}\right]\right)$$

the density of $(Y_i)_{i=1}^K$ over p, or equivalently, minimizing $\Sigma[(Y_i - p)^2/\sigma_i^2]$ as a function of p. By differentiation, we get

$$\hat{p} = \Sigma(Y_i/\sigma_i^2)/\Sigma(1/\sigma_i^2)$$

as the estimate and

$$\Sigma[(Y_i - \hat{p})^2/\sigma_i^2]$$

as the minimizing value.

The likelihood ratio test rejects the null hypothesis when the ratio of the density of $(Y_i, \ldots, Y_K)$ evaluated at $\hat{p} = Y_i$; (the maximum likelihood estimate of p_i under the alternative hypothesis) to the density of $(Y_i, \ldots, Y_K)$ evaluated at $\hat{p} = \Sigma(Y_i/\sigma_i^2)/\Sigma(1/\sigma_i^2)$ (the maximum likelihood estimate under the null hypothesis) is large. Referring to the previous paragraph, we see that the numerator reduces to a constant while the denominator is exp $[-\frac{1}{2}\Sigma(Y_i - \hat{p})^2/\sigma_i^2]$ so it is appropriate to reject when $[\Sigma(Y_i - \hat{p})^2]/\sigma_i^2$ is large, and we replace $\sigma_i^2 = p_i(1-p_i)/n_i$ by $\hat{p}(1-\hat{p})/n_i$ in accordance with the above discussion.

Then

$$\hat{p} = \frac{\sum\frac{Y_i}{\sigma_i^2}}{\sum\frac{1}{\sigma_i^2}} = \frac{\sum\left[\frac{X_i}{n_i}\frac{1}{p(1-p/)n_i}\right]}{\sum\frac{1}{p(1-p)/n_i}} = \frac{\Sigma X_i}{\Sigma n_i}$$

and

$$\sum\frac{Y_i - \hat{p})^2}{\sigma_i^2} = \frac{\sum\left(\frac{X_i}{n_i} - \frac{\Sigma X_i}{\Sigma n_i}\right)^2}{\hat{p}(1-\hat{p})/n_i}$$

$$= \frac{1}{\hat{p}(1-\hat{p})}\sum n_i\left(\frac{X_i}{n_i} - \frac{\Sigma X_i}{\Sigma n_i}\right)^2$$

Denoting p_i as X_i/n_i (the estimate of the p_i of the jth binomial), then we reject H_0 when:

$$\frac{1}{\hat{p}(1-\hat{p})}\Sigma n_i\left(\hat{p}_i - \frac{\Sigma n_i\hat{p}_i}{\Sigma n_i}\right)^2$$

is large.

Where $\Sigma n_i\hat{p}_i/\Sigma n_i = \Sigma X_i/n_i = \hat{p}$ is the estimate of p if all p_i's are equal. Equivalently, we reject H_0 if $1/[p(1-\hat{p})]\Sigma n_i(\hat{p}_s - \hat{p})^2$ is large with p_i and p defined as above.

A Method for Testing the Hypothesis under Assumption II:

X_i is Poisson with parameter $n_ip_i = \lambda_i$.

R. A. Fisher proved that the conditional distribution of X_i given ΣX_i is multinomial, with parameters $\lambda_i/\Sigma\lambda_i$ and ΣX_i.

Under the null hypothesis that $p_i = p_j$ for all i and j, then $\lambda_i/\Sigma\lambda_i = n_i/\Sigma n_i$. For testing hypotheses of this type involving Poissons, certain desirable properties are maintained when the problem is treated as a multinomial (obtained by Fisher's theorem) and if one uses an appropriate multinomial test. The test which most immediately suggests itself is the χ^2 test (chi square test).

In the notation of the preceding section, $\hat{p}_i = X_i/n_i$ and $\hat{p} = \Sigma X_i/\Sigma n_i$ so the χ^2 test is to reject for large values of $(1/\hat{p})\Sigma n_i(\hat{p}_i - \hat{p})^2$. This test differs from the previous by a factor of $1/(1-\hat{p})$, and hence by very little since the Poisson assumption is only valid when $\hat{p}$ is near zero and therefore when $1/(1-\hat{p})$ is near unity. The approximate distribution of the statistic is in any case $\chi^2_{(K-1)}$, the χ^2 statistic with $K-1$ degrees of freedom.

It is a curious fact that the statistic $[1/(1-\hat{p})(\hat{p})]\Sigma n_i(\hat{p}_i - \hat{p})^2$ is exactly equal to the χ^2 statistic suggested by Karl Pearson for testing the hypothesis of independence

in a $2 \times K$ contingency table when the marginal totals are regarded as fixed.

Simultaneously, the hypothesis of independence in a $2 \times K$ contingency table places precisely the same restrictions on the parameters p_{ij}, $j = 1, 2$; $i = 1, \ldots, K$ as the hypothesis of equality of means for independent binomials places on the $(p_{ij}\ 1 - p_i)$ $i = 1, \ldots, K$, *so the hypotheses and test statistics are identical.*

Problem 2

Given: A two-way contingency table, to test that there is a correlation between the factors, i.e.,

H_0: no correlation
H_1: there is a correlation

	Not B	B	
Not A	n_{00}	n_{01}	$n_{0\cdot}$ = total
A	n_{10}	n_{11}	$n_{1\cdot}$ = total
	$n_{\cdot 0}$	$n_{\cdot 1}$	$n_{\cdot\cdot}$ = grand total

Y

X	0	1	
0	p_{00}	p_{01}	$p_{0\cdot}$
1	p_{10}	p_{11}	$p_{1\cdot}$
	$p_{\cdot 0}$	$p_{\cdot 1}$	1

The population correlation is defined as

$$r = \frac{E \times Y - E \times EY}{(\text{var } X \cdot \text{var } Y)^{1/2}}$$

$$= \frac{P_{11} - P_{1\cdot}P_{\cdot 1}}{(\text{var } X \cdot \text{var } Y)^{1/2}}$$

where $X = 1$ if A occurs, 0 if not, $Y = 1$ if B occurs, 0 if not. Hence $\pi = 0 \Rightarrow p_{11} = p_{1\cdot}p_{\cdot 1}$ which is the hypothesis of independence: thus we may test H_0 by testing independence. But then the χ^2 test is used with one degree of freedom:

$$\chi^2 = \sum_{j=1}^{2}\sum_{i=1}^{2} \frac{(0_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \sum_{j=1}^{2}\sum_{i=1}^{2} \times \left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n_{\cdot\cdot}}\right)^2 \Big/ \frac{n_{i\cdot}n_{\cdot j}}{n_{\cdot\cdot}}$$

Here $E_{ij} = n_{\cdot\cdot}(n_{i\cdot}/n_{\cdot\cdot})(n_{\cdot j}/n_{\cdot\cdot})$ where $n_{i\cdot}/n_{\cdot\cdot}$ is the probability that A falls in the ith cell when the marginal totals are fixed. If the $n_{i\cdot}$ are not fixed, then there is a small error due to using this test, the error being unimportant if n_i and n_j are very large, as they are in this case. Or, alternatively, one may condition on the values of the $n_{i\cdot}$ and $n_{\cdot j}$ and regard the test as a conditional test of hypothesis. In any case, the test may be used in either situation.

Another way to interpret the above problem is to suppose that one wishes to compare $p = P(X = 1 \mid Y = 1) = P(A \mid B)$ with $P(A)$. [(Positive correlation would make $P(A \mid B) > P(A)$].

Then in some intuitive sense one would have some feeling about the relation between A and B that might not be provided by a correlation coefficient. If A and B are independent, however, $P(A|B) = P(A)$. The estimate for $P(A|B)$ = the fraction of times A occurs, given B occurs, and is then $n_{11}/n_{\cdot 1}$.

Suppose $n_{\cdot 1}$ to be fixed, the n_{11} is binomial $b(p,n_{\cdot 1})$ and therefore with $\hat{p} = n_{11}/n_{\cdot 1}$, $E\hat{p} = p$, $\text{Var}\,\hat{p} = 1/n_{1\cdot}p(1-p) \leq 1/4n_{1\cdot}$ since the maximum is attained at $p = {}^1/_2$. Finally, by the normal approximation to the binomial, $\hat{p} \sim n(p, {}^1/_4 n_{1\cdot})$ where the variance is not given exactly, but is conservative. But then a confidence interval may be obtained for r by the usual methods: for example, if

$$p\left(-\frac{K}{(4n_{1\cdot})} \leq \hat{p} - p \leq \frac{K}{(4n_{1\cdot})^{1/2}}\right) = 0.95$$

then k is the 97.5% cutoff for a normal ($k = 1.96$). So if we regard k as 2:

$$p\left(-\frac{2}{(4n_{1\cdot})^{1/2}} \leq \hat{p} - p \leq \frac{2}{(4n_{1\cdot})^{1/2}}\right)$$

$$= \left(p \text{ lies in } \hat{p} \pm \frac{1}{(n_{1\cdot})^{1/2}} = 0.95\right)$$

This set of values for p may be compared to $P(A) = n_{1\cdot}/n_{11\cdot} = p_{1\cdot}$ to determine whether or not $p = P(A/B)$ and $P(A)$ differ significantly.